



Unveiling precision: a data-driven approach to enhance photoacoustic imaging with sparse data

MENGYUAN HUANG,¹ WU LIU,¹ GUOCHENG SUN,¹ CHAOJING SHI,¹ XI LIU,¹ KAITAI HAN,¹ SHITOU LIU,¹  ZIJUN WANG,¹ ZHENNIAN XIE,^{2,4} AND QIANJIN GUO^{1,3,5,6} 

¹Academy of Artificial Intelligence, Beijing Institute of Petrochemical Technology, Beijing 102617, China

²Xiyuan Hospital, Chinese Academy of Traditional Chinese Medicine, China

³School of Mechanical Engineering & Hydrogen Energy Research Centre, Beijing Institute of Petrochemical Technology, Beijing 102617, China

⁴xiezhennian@163.com

⁵guoqianjin@bipt.edu.cn

⁶guojq@iccas.ac.cn

Abstract: This study presents the Fourier Decay Perception Generative Adversarial Network (FDP-GAN), an innovative approach dedicated to alleviating limitations in photoacoustic imaging stemming from restricted sensor availability and biological tissue heterogeneity. By integrating diverse photoacoustic data, FDP-GAN notably enhances image fidelity and reduces artifacts, particularly in scenarios of low sampling. Its demonstrated effectiveness highlights its potential for substantial contributions to clinical applications, marking a significant stride in addressing pertinent challenges within the realm of photoacoustic acquisition techniques.

© 2023 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

The field of bioimaging has seen increasing use of photoacoustic imaging, a non-invasive hybrid imaging method, thanks to its swift advancement in recent years. Photoacoustic imaging (PAI) encompasses two forms of imaging: Photoacoustic Microscopy (PAM) and Photoacoustic Tomography or Computed Tomography (PAT/PACT) [1]. PAM provides high-resolution images of superficial tissues, whereas PAT is optimal for deep tissue exploration. As a result, PAT has broad application prospects in various biomedical imaging diagnostic fields, such as early cancer detection [2], cellular imaging [3], biopsy guidance [4], and more.

Practical implementation of PAI often encounters challenges due to less-than-ideal sampling of tomographic data. Acquiring high-quality images usually necessitates a higher number of transducers during data acquisition, but this solution is accompanied by higher costs, escalated hardware complexity, and a demand for greater computational power. Consequently, a more feasible approach limits the number of transducers and conducts measurements under sparse viewing conditions. However, sparse-view photoacoustic imaging, while reducing system intricacy and cost, may result in compromised image quality, exhibiting artifacts and blurring. Complications also arise during live imaging, as the movement of the organism can introduce image artifacts, thus undermining the image quality. These factors play a crucial role in enhancing the accuracy and sensitivity of neuronal activity detection [5,6]. In multispectral photoacoustic tomography, changes in excitation wavelengths at a given location may result in motion artifacts, affecting the resolution of photoacoustic results and quantitative analysis. Additionally, neglecting acoustic heterogeneity can have negative consequences on the resolution and image quality unless properly mitigated using appropriate reconstruction algorithms [7–9]. The constraints of transducer scan range and sparse viewing angle during sampling typically result in compromised

image acquisition quality [6,10]. These challenges drive the research community's ongoing efforts to enhance the resolution, speed, sensitivity, and depth of photoacoustic and tomographic imaging. The critical issue of noise reduction and artifact elimination in photoacoustic images derived from limited and sparse views continues to be a key focus in contemporary research.

In recent decades, artificial intelligence (AI) has become a research hotspot, and its application in medical imaging, particularly deep learning, has gained significant attention from researchers and clinicians. Photoacoustic imaging has also benefited from advances in AI, especially in PAI reconstruction technology. Efficient reconstruction methods based on deep learning, such as U-Net, Generative Adversarial Networks (GAN), and Transformers, have been successfully applied by researchers in medical image processing [11–14]. For instance, Neda Davoudi et al. [10] proposed a novel framework for deep convolutional neural networks to recover image quality from sparse photoacoustic data, eliminating some reconstruction artifacts present in sparsely sampled data. This method accelerates data acquisition and image reconstruction processes by reducing common image artifacts, enhancing anatomical contrast, and improving image quantization capabilities. Stephan Antholzer et al. [13] developed a straightforward and efficient reconstruction algorithm based on deep learning. In this method, image reconstruction is performed using a deep convolutional neural network (CNN) whose weights are adjusted prior to the actual image reconstruction based on a set of training data. The proposed reconstruction method can be interpreted as a network which uses PAT filtered inverse projection algorithm in the first layer and then U-net architecture in the remaining layers. Real image reconstruction using deep learning requires only one evaluation of the trained CNN and does not require time-consuming solving of the forward and concomitant problems. Steven Guan et al. [15] proposed an improved CNN architecture called FD-UNet for removing artifacts from 2D PAT images reconstructed from sparse data. FD-UNet and UNet were compared by using datasets generated from synthetic models (circles, Shepp-Logan, and vasculature system) and anatomically realistic datasets of the mouse brain vasculature system. FD-UNet proved to be the superior and more compact CNN for removing artifacts and improving image quality.

Although systematic studies investigating the performance of deep learning methodologies with varying detection angles at finite viewpoints exist, the efficacy of deep learning techniques employing diverse numbers of transducers at sparse-view photoacoustic imaging (PA) remains under-explored. In this context, we introduce an enhanced GAN methodology, the FDP-GAN, and examine its efficacy for image reconstruction from various datasets under restricted viewing angles. This work's primary contributions include:

1. The FDP-GAN model, which is based on GAN architecture, undergoes training using a cost function that incorporates regularization. Its distinctive feature is its capability to restore image quality even with a limited number of transducers. To our knowledge, this is the first instance of PA image quality restoration using such limited transducers and manually added noise, which simultaneously reduces the technique's cost and increases its clinical utility.
2. We amplified the precision and generalizability of FDP-GAN by training and testing the network on simulated and publicly available datasets. This approach mitigated the requirement for substantial experimental training datasets, hence enhancing FDP-GAN's applicability in diverse scenarios. We also utilized three conventional metrics to quantitatively evaluate FDP-GAN's performance in different limited view situations featuring representative structures.
3. We manually acquired photoacoustic data using the OR-PAM device and conducted experiments on the PAM and PAT datasets to improve image quality from sparse data, revealing that FDP-GAN can recover PA images without additional training, thus demonstrating its versatility and broad applicability. In addition, the ability of FDP-GAN to run without

retraining saves computational resources and data storage space, making it suitable for practical clinical applications.

The paper's structure is as follows: Section 2 elaborates on the methodology employed to create and construct the simulation datasets, and delves into the network structure model utilized. Section 3 describes the detailed procedure of this experiment and uses metrics to evaluate the image reconstruction results. The results and discussion of this work are presented in Sections 4 and 5, respectively.

2. Materials and methods

2.1. Datasets preparation

This study employed two data acquisition methods. Firstly, optical-resolution photoacoustic microscopy (OR-PAM) was utilized to acquire photoacoustic data. These data were derived from dragonfly wings [16,17] and magpie feathers [18–20], totaling approximately 1023 images, comprising 203 high-quality, clear images, and 820 under-sampled and sparse images. These data were employed for transfer learning of the model proposed in this study to assess its performance. The OR-PAM system was configured as depicted in Fig. 1(ii), employing a pulsed laser with a 532 nm wavelength (VPFL-G-30, Spectra-Physics) featuring a pulse width of 7 ns and a pulse energy of approximately 80 nJ. Initially, the laser beam was transmitted through a 2-meter single-mode fiber (SMF, P1-460B-FC-2, Thorlabs Inc), with a coupling efficiency of approximately 60%, to the PAM probe. Subsequently, the beam combiner in the PAM probe reflected the light onto the sample while transmitting the generated ultrasonic waves to a piezoelectric transducer (V214-BC-RM, Olympus-NDT). As the light passed through a polarization beam splitter (PBS, PBS051, Thorlabs Inc) and was transmitted to the SMF, it was tightly focused on the tissue sample. The lateral resolution of the OR-PAM module in our system was approximately 4.5 μm .

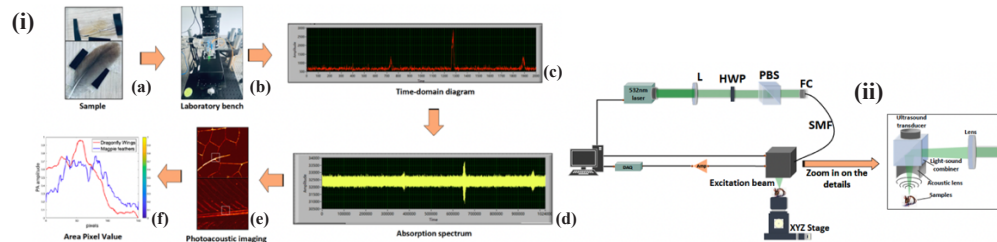


Fig. 1. Experimental System Diagram. Figure 1(i) delineates the comprehensive flowchart of the experimental system, while Fig. 1(ii) provides a detailed exposition of the optical path diagram employed in the experiment.

Our experimental procedure, as illustrated in Fig. 1(i), commences by situating samples collected from natural environments onto the experimental platform. Subsequently, we initiate photoacoustic wave generation through laser excitation, driven by tissue's optical absorption characteristics. The experimental setup, as delineated in Fig. 1(ii), encompasses pivotal components for laser emission, photoacoustic wave generation, detection, data acquisition, and subsequent computational processing. These resultant photoacoustic waves are then systematically detected and imaged using ultrasound methodologies, enabling the extraction of pertinent structural and functional information concerning the tissue. Temporal profiles corresponding to individual stages of sample scanning are elucidated in Fig. 1(i-c), with corresponding absorption spectra featured in Fig. 1(i-d). Photoacoustic images (Fig. 1(i-e)) provide visual insights into tissue morphology, while the pixel values (Fig. 1(i-f)) facilitate quantitative analyse. After screening the acquired images, the processed set of datasets was

divided into two subsets. One subset was dedicated to training the model using diverse deep learning algorithms, while the other subset served as a test set to assess the model's performance.

Secondly, this work also uses a publicly available dataset to validate the effectiveness of the method on PAT image data from phantom imaging experiments performed by Neda Davoudi et al. [10]. This publicly available datasets from <https://doi.org/10.6084/m9.figshare.9250784>. Phantom training data were generated by imaging paper-printed absorptive targets embedded in 16 mm diameter agar cylinders. Agar substrates were prepared by diluting 1.3% (w/w) agar powder in water. A total of 28 phantoms were measured, from which we collected 450 images by data enhancement with random rotation and panning. For each phantom, the printing paper contained three randomly distributed 2.5-mm-diameter black circles. Thirty-three phantoms consisting of printed vascular structures adapted from images obtained using an OA microscope setup were also measured, from which 100 images were collected by data enhancement with random rotation. These images were acquired utilizing an 80 mm diameter, 512-element ring-shaped detector array. These two data acquisition methods provide diverse data resources for transfer learning and performance evaluation of the model in this study.

2.2. Framework of the FDP-GAN model

To enhance image details and address the smoothing tendency of traditional convolutional networks in insufficient sampling conditions, this work use 16-256 transducers for image enhancement. The meticulous selection of this specific sensor range is underpinned by its profound impact on bolstering image fidelity and precision.

The FDP-GAN Model adeptly mitigates the smoothing tendency in traditional convolutional networks under limited sampling. We added Fourier transform, decay function, perceptual loss function, and L_1 loss to the model to achieve better results in contour edges. This study is centered on evaluating the efficacy of FDP-GAN networks when operating under conditions characterized by limited sampling, specifically employing a configuration involving only a low number of sensor configurations. The FDP-GAN network, developed in this study, was specifically tailored for the mice dataset gathered during the PAT experiment. Its impressive level of generalization was thoroughly validated on the PAM dataset.

The framework of the FDP-GAN network architecture model is illustrated in Fig. 2. The process can be quantitatively expressed as:

$$\mathcal{L}_{GAN}(G, D) = E_y[\log D(y)] + E_{x,z}[\log(1 - D(G(x, z)))] \quad (1)$$

To address the unsuitability of unsupervised GAN networks, which are based on the random generation of noise, for paired low and high sampling datasets, we have employed a conditional adversarial generative network (CGAN) based on a Pix2Pix model as the underlying backbone network. The process is illustrated in Eq. (2). This approach enhances the robustness and stability of the generation process, ensuring reliable results.

$$\mathcal{L}_{cGAN}(G, D) = E_{x,y}[\log D(x, y)] + E_{x,z}[\log(1 - D(x, G(x, z)))] \quad (2)$$

This work demonstrates that the FDP-GAN model is effective in denoising PA data. [21] Unlike traditional GAN models, the Pix2Pix GAN model requires paired data for training. Specifically, the discriminator is trained on both $\langle x, y \rangle$ and $\langle x, G(x) \rangle$ pairs, where $G(x)$ represents the output image generated by the generator from the input image x . The objective of training the Pix2Pix GAN model is to minimize the following loss function:

$$\mathcal{L}_{L_1}(G) = E_{x,y,z}[\|y - G(x, z)\|_1] \quad (3)$$

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L_1}(G) \quad (4)$$

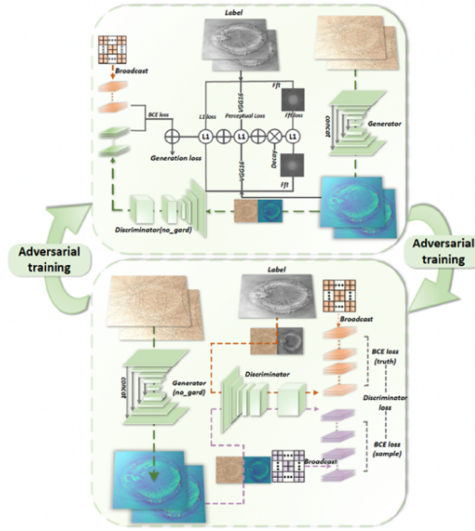


Fig. 2. Structure of the network of FDP-GAN.

where the terms \mathcal{L}_{cGAN} and \mathcal{L}_{L_1} represent the CGAN loss and L_1 distance, respectively. The expected value is denoted by E , where z represents a random noise vector of the same shape as x . The real sample is represented by y , while G and D refer to the generator and discriminator networks, respectively.

Emulating a classical and proficient discriminator architecture, the FDP-GAN network incorporates a structure with five hidden layers. The image passes through the discriminator to obtain a patch of a 32×64 matrix, and the closer the matrix is to the 1 matrices of 32×64 , the more the generated image can confuse the discriminator, indicating that the generated image is similar to the real sample. Conversely, the closer the matrix is to the 0 matrix of 32×64 , the more likely the discriminator is to judge the generated image as a non-real sample.

In the network structure design, we use the “broadcast” mechanism, which is a mechanism for handling mathematical operations between tensors of different shapes. It allows arithmetic operations to be performed by automatically expanding tensors with different shapes so that they have the same shape, allowing element-by-element operations to be performed. The benefit of this mechanism is that it allows us to save memory and computational resources by avoiding explicit dimension expansion or replication when dealing with data of different shapes. Thus, the loss function of our proposed FDP-GAN model is carefully designed as follows:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [\alpha * L_1 Loss + BCE Loss + \gamma * FFT Loss + \beta * Perceptual Loss] \quad (5)$$

Figure 3 provides a detailed illustration. This approach ensures that the loss function can directly align with the criteria of the input and target images. The network model was developed using Python 3.9 and PyTorch-v2.0, an open-source deep-learning library. The network training and evaluation were conducted on an NVIDIA GTX 2080Ti GPU. During the training process, the mean square error loss was minimized using the Adam optimizer. The training utilized a batch size of 2 and spanned 300 epochs, with an initial learning rate of $2e^{-4}$. During training, the FDP-GAN model was provided with paired training datasets $\{x^i, y^i\}$, where x^i represents the input data, and y^i represents the corresponding 512 sensor images. The specific pseudo-code for training can be observed in Algorithm 1 and Algorithm 2:

Algorithm 1. Generator Training

The generator uses Unet with residual connections as the backbone network, and uses Fourier loss, perceptual loss, and L_1 loss as gradient descent targets. Parameter j is the feature representation of the J layer in the pretrained neural network, and N is the number of feature layers. α, β is an adjustable hyperparameter. γ is a parameter of decay. We use exponential decay at the base of the natural logarithm with a half-life of 100 epochs.

For number of train iterations **do**

- Sample batch of m samples $\{x^{(1)}, \dots, x^{(m)}\}$ from low number of projections
- Sample batch of m labels $\{y^{(1)}, \dots, y^{(m)}\}$ from 512 number of projections
- Update the generator by descending its stochastic gradients:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [\alpha * L_1Loss + BCELoss + \gamma * FFTLoss + \beta * PerceptualLoss]$$

1. Loss Functions

$$L_1Loss = |y^{(i)} - G(x^{(i)})|$$

$$BCELoss = -1 \log(D(G(x^{(i)})))$$

$$FFTLoss = |fft(\log(G(x^{(i)}))) - fft(\log(y^{(i)}))|$$

$$PerceptualLoss = \frac{1}{N} \sum_{j=1}^N (F_i(G(x^{(i)})) - F_i(y^{(i)}))^2$$

2. Decay Function

$$\gamma = e^{\frac{-x \ln 2}{100}}$$

$$x = \text{epoch nums}$$

End for

The gradient-based updates can be implemented using various standard gradient-based learning rules. In this particular experiment, we employed the momentum method.

Algorithm 2. Discriminator Training

Discriminator uses the PatchGAN structure, PatchGAN is designed in the form of complete convolution and uses convolution to map input into $N \times N$ matrix to improve discriminator receptive field, so that more areas can be focused.

For the number of train iterations **do**

- ◆ Sample batch of m samples $\{x^{(1)}, \dots, x^{(m)}\}$ from low number of projections
- ◆ Sample batch of m labels $\{y^{(1)}, \dots, y^{(m)}\}$ from 512 number of projections
- ◆ Update the discriminator by ascending its stochastic gradients:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [-\log(1 - D(x^{(i)})) - \log(D(y^{(i)}))]$$

End for

The gradient-based updates can employ any standard gradient-based learning rule. In this experiment, we utilized the momentum method.

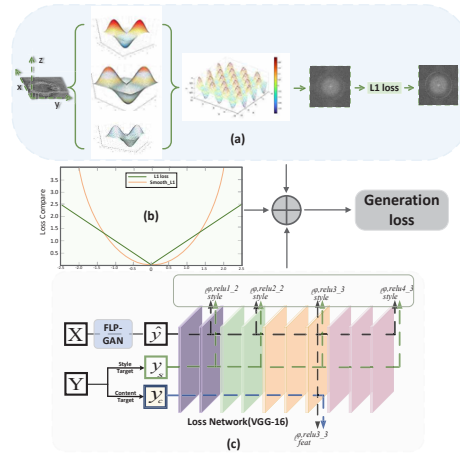


Fig. 3. The details of the proposed FDP-GAN model.

2.2.1. FFT-based loss

In the task of image recovery, it is crucial to capture both low and high-frequency information. However, conventional Residual Blocks have limitations in incorporating high-frequency features. To address this issue, we put forth the FFT+ L_1 configuration for evaluating loss, as illustrated in Fig. 3. This structure incorporates two distinct branches: the time-honored L_1 loss branch and the innovative FFT branch.

Here, as illustrated in Fig. 3(a), the original input image undergoes a discrete Fourier transform followed by spectral centering of the resulting spectral map. Most of the images are represented as grayscale maps, where each pixel is assigned a grayscale value.

In this work, we mainly use the Fourier transform without the inverse transform to directly approach the images in the frequency domain. The generated and labeled maps are computed directly in the frequency domain space to obtain the L_1 loss, which represents their distance in the frequency domain space. The information reconstruction in the frequency domain is controlled by gradient descent:

$$\text{FFTLoss} = |\text{fft}(\log(G(x^{(i)}))) - \text{fft}(\log(y^{(i)}))| \quad (6)$$

where $x^{(i)}$ represents a batch of m samples randomly selected from a limited set of projections $\{x^{(1)}, \dots, x^{(m)}\}$, while $y^{(i)}$ represents a batch of m labels selected from a larger set of 512 projections $\{y^{(1)}, \dots, y^{(m)}\}$.

2.2.2. Decay function

In the current research, decay functions [22,23] have become widely used to solve physical problems. To compensate for some of the lost image edge information and highlight the edge contours, we introduced the decay function into the training model [24]. By incorporating a decay function [25] in the training model, we introduce a mechanism to mitigate this smoothing effect. The decay function effectively controls the rate of decay at certain frequencies, allowing us to prioritize the retention of high-frequency components associated with image edges.

In essence, the attenuation function acts as a correction to ensure that critical edge information is preserved and faithfully rendered in the reconstructed image. This is especially important for accurate reconstruction of fine details and structural features when data is sparse or sampling is limited. Decay function limits the overall training loss value to below 0.1. This function is

defined as follows:

$$\gamma = e^{\frac{-x \ln 2}{100}} \quad (7)$$

where x stands for the epoch nums and the weight decay function employed in the Fourier transform experiments is represented by the parameter γ . We utilized exponential decay with a fixed half-life of 100 epochs to implement this function. Additional information on the specific experimental setup can be found in Section 3.2.

2.2.3. Perceptual loss function

To rectify the issues of sparse views and indistinct edges encountered, we've implemented the perceptual loss function put forth by Johnson et al. [26] The fixed network harnesses both ground truth images and the network-produced results as inputs, facilitating the generation of output features.

We employed the VGG16 module (depicted in Fig. 3(c)) for feature extraction in our FDP-GAN architecture. The perceptual loss was derived from the output of four activation layers (highlighted in red boxes), corresponding to sections outlined in blue in the FDP-GAN structure diagram. The construction of the perceptual loss involved setting up a static network using a pre-trained VGG16 model on ImageNet, with parameters held constant during training. Subsequently, the actual image (referred to as Ground Truth) and the network's output (referred to as Prediction) were input into the static network, producing output features designated as $feature_gt$ and $feature_pre$. The final step involved calculating the L_2 distance between $feature_gt$ and $feature_pre$ to generate a loss. This loss acts as an approximation of the depth information disparity between the actual image and the network's output, thereby enhancing detailed information beyond the typical L_2 loss. Two key points should be noted in this process:

1. Loss Composition: The loss function of a generative network typically encapsulates a blend of multiple losses, which together define the final loss function. This can be represented as:

$$L_{total} = \alpha L_{rec} + \beta L_{per} \quad (8)$$

where α and β are used as trade-off coefficients to adjust the importance of different losses to the total loss function.

2. Perceptual Enhancement: The application of perceptual losses enhances the granularity of the generated image, surpassing the outcomes solely driven by L_2 loss. Given the real image y and the generated image, the loss function can be expressed as follows:

$$y = (y_1, y_2, \dots, y_N) \bar{y} = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_N) \quad (9)$$

Here, N corresponds to the sample size within a batch. The term N denotes the N th network layer, symbolizing the output features of that respective layer.

3. Experiments and results

3.1. FDP-GAN method

3.1.1. Experimental setup

In this work, for the sake of consistent comparison, certain experimental parameters were uniformly applied across different models. All models were created, trained, and evaluated using the Pytorch platform with an NVIDIA GeForce RTX 2080Ti graphics card for computation. During the training phase, we set the neural network's momentum at 0.9, the batch size at 8, and the learning rate at 0.0001. Also, we allowed a maximum of 300 epochs for the model comparison. After conducting several experiments, we discovered that an initial learning rate of 2×10^{-4} was optimal.

3.1.2. Data preprocessing

To ensure network convergence and facilitate seamless learning, it is imperative to apply data normalization techniques, particularly when variations in distribution or scale of identical modal features are anticipated. The images in the dataset were divided into a training set, comprising 80% of the data, and a test set, comprising the remaining 20%. The mean square loss is monitored during validation. To prepare the training set, the artifact-damaged image $X \in R^{N \times N}$ and the artifact-free image $Z \in R^{N \times N}$ are paired together as $T = (X_m, Z_m)_{m=1}^M$, following standard neural network theory. During the optimization process, the mapping function is determined based on the training error:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [\alpha * L_1 Loss + BCE Loss + \gamma * FFT Loss + \beta * Perceptual Loss] \quad (10)$$

The main goal of training neural networks is to minimize the aforementioned equations, consequently lowering the training error or loss. However, to construct neural networks that have a more direct correlation with the mapping function and can be designed in a systematically manner, the establishment of a distinct and comprehensively defined framework is of utmost importance:

$$A_w = (\sigma_L \circ W_L) \circ \dots \circ (\sigma_1 \circ W_1) \quad (11)$$

where σ represents the activation function, W denotes the weighting vector, and L corresponds to the number of subsequent processing layers. The weighting vector is a critical parameter in network optimization and is continually updated based on the reference image.

3.1.3. Evaluation metrics

To evaluate the performance of the FDP-GAN model, two quantitative measures were employed. These included the widely recognized Structural Similarity Index (SSIM), which amalgamates statistical measurements like mean, variance, covariance, and correlation, and the Peak Signal-to-Noise Ratio (PSNR), both of which are classical criteria for evaluating image quality. The PSNR offers a general quality assessment of the image, whereas the SSIM delivers local measurements, taking into consideration elements like contrast, luminance, and structural resemblance.

$$PSNR = 20 \log_{10}(255/RMSE) \quad (12)$$

$$RMSE = \sqrt{\frac{1}{pq} \|X - X_{ori}\|_F^2} \quad (13)$$

where p refers to the row count of the image, q corresponds to the column count of the image, X stands for the image after reconstruction, and X_{ori} is the benchmark or reference image. SSIM is a measure designed to evaluate the visual impact of three characteristics of an image: luminance, contrast, and structure. This perceptual metric identifies image quality degradation, with structural information being described as interdependent pixels that are spatially close.

$$SSIM(x, y) = [l(x, y)]^a \cdot [c(x, y)]^b \cdot [s(x, y)]^g \quad (14)$$

The above parameters hinge on three different elements: l , which compares the brightness of the predicted and original image; c , representing contrast; and s , comparing the structure of the two images. Furthermore, a , b , and g are positive constants, while x and y correspond to the original and reconstructed images. The brightness, contrast, and structure are influenced by

specific factors:

$$\text{brightness: } l(X, X_{ori}) = \frac{2\mu_x\mu_{x_{ori}} + C_1}{\mu_x^2 + \mu_{x_{ori}}^2 + C_1} \quad (15)$$

$$\text{contrast: } c(X, X_{ori}) = \frac{2\sigma_x\sigma_{x_{ori}} + C_2}{\sigma_x^2 + \sigma_{x_{ori}}^2 + C_2} \quad (16)$$

$$\text{structure: } s(X, X_{ori}) = \frac{2\sigma_{xx_{ori}} + C_3}{\sigma_x\sigma_{x_{ori}} + C_3} \quad (17)$$

$$\sigma_{xx_{ori}} = \frac{1}{N-1} \sum_{j=1}^N (X_j - \mu_x)((X_{ori})_j - \mu_{x_{ori}}) \quad (18)$$

where α_1 , α_2 , and α_3 are weights, X_i denotes the i th pixel value that is used to construct the image vector X , μ_x is the luminance estimate of X , σ_x is the contrast estimate of X , and C_1 , C_2 , and C_3 are constants. In this experiments, we set α_1 , α_2 , and α_3 to 1, and C_1 , C_2 , and C_3 to $C_1 = (K_1L)^2$, $C_2 = (K_2L)^2$, $C_3 = C_2/2$. The maximum pixel value of the image, denoted by L , is typically 255 for grayscale images. By assigning small constants K_1 and K_2 , specifically $k_1 = 0.05$ and $k_2 = 0.05$. Enhanced reconstruction quality and accurate structure can be assessed by analyzing the PSNR and SSIM, with higher values indicating better results.

3.2. Experimental result

3.2.1. Performance evaluation on PAM datasets

In the process of photoacoustic imaging, tissue heterogeneity leads to acoustic reflection and imaging artifacts, which subsequently affect the imaging outcome. Acoustic reflection results in the return of sound waves from the internal tissue, potentially interfering with clear imaging in the target region. Concurrently, heterogeneity induces scattering of the sound waves within the tissue, giving rise to artifacts in the image and diminishing the accuracy and clarity of the resulting visualization. Therefore, a thorough understanding and management of tissue heterogeneity is crucial in photoacoustic imaging.

We substantiate the FDP-GAN model's generalizability through the utilization of manually acquired OR-PAM data. The experimentally acquired image is depicted in Fig. 4, revealing evident noise and a deficiency in edge details. We employed the FDP-GAN network structure to reconstruct the experimental data. A comparative analysis of three specific regions, z1, z2, and z3, illustrates marked improvements in image detail and sharpness. The quantitative evaluation results of the images are presented in Table 1.

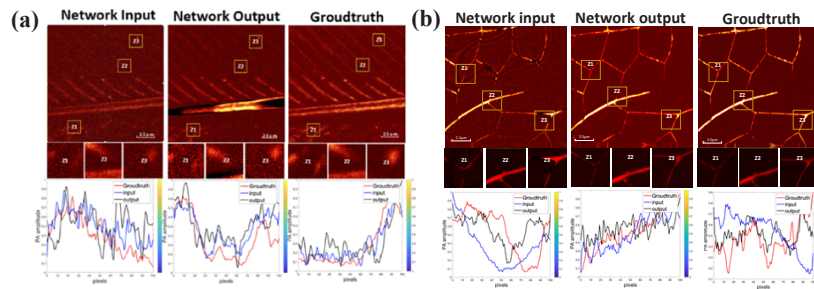


Fig. 4. PAM vascular images of feathers and dragonflies wings, acquired using the OR-PAM experimental apparatus, were subjected to imaging transformation through the utilization of the pre-trained FDP-GAN model. In Figs (a) and (b), the left column shows a zoomed-in OR-PAM image. The middle column shows the image obtained after network training and the right column shows the standard ground truth.

Table 1. List of comparisons of different networks for different samples^a

Samples	Raw	CNN	UNet	ResCBAM	TransUnet	ResUnet	pix2pix	FDP-GAN
Feathers-ssim	0.515	0.602	0.661	0.674	0.709	0.667	0.742	0.773
Wings-ssim	0.445	0.526	0.551	0.601	0.676	0.622	0.707	0.739
Feathers-psnr	22.114	24.218	25.301	26.664	28.783	26.997	29.890	30.565
Wings-psnr	20.217	22.546	24.003	25.476	27.063	25.818	28.156	29.390

^aSSIM = structural similarity index measure; PSNR = peak signal-to-noise ratio.

Table 2. Comparison of SSIM metrics for all three networks by using 256-, 128-, 64-, 32-, 16-, and 8-projection information^a

Network used	8-Projection	16-Projection	32-Projection	64-Projection	128-Projection	256-Projection
Raw	0.252	0.275	0.461	0.531	0.619	0.827
CNN(3 layers)	0.385	0.446	0.501	0.686	0.806	0.881
UNet(pure)	0.503	0.558	0.688	0.762	0.842	0.931
ResCBAM	0.530	0.566	0.710	0.808	0.889	0.950
TransUnet	0.536	0.571	0.707	0.770	0.849	0.938
ResUnet	0.537	0.572	0.709	0.775	0.848	0.946
pix2pix	0.571	0.638	0.752	0.809	0.890	0.950
FDP-GAN	0.593	0.749	0.754	0.816	0.893	0.948

^aSSIM = structural similarity index measure.

Table 3. Comparison of PSNR metrics for all three networks by using 256-, 128-, 64-, 32-, 16-, and 8-projection information^a

Network used	8-Projection	16-Projection	32-Projection	64-Projection	128-Projection	256-Projection
Raw	17.563	17.825	19.058	22.323	23.268	31.188
CNN(3 layers)	18.127	18.873	20.144	25.369	30.874	32.116
UNet(pure)	20.652	22.550	26.037	29.365	31.384	32.908
ResCBAM	22.125	22.696	27.812	30.649	32.194	33.936
TransUnet	22.334	22.877	27.224	30.168	31.547	33.193
ResUnet	22.372	22.899	27.689	30.223	31.960	33.398
pix2pix	23.373	25.725	28.284	30.675	32.213	33.936
FDP-GAN	23.389	27.113	28.296	30.696	32.343	33.939

^aPSNR = peak signal-to-noise ratio.

Table 4. Quantification comparison of ablation experiments (mean \pm std)^a

Methods	SSIM(16 projections)	PSNR(16 projections)
Only decay	0.570 \pm 0.05	22.80 \pm 0.05db
Decay and fourier loss	0.610 \pm 0.06	23.90 \pm 0.05db
Decay and perceptual loss	0.630 \pm 0.04	24.40 \pm 0.07db
Decay ,fourier loss and perceptual loss	0.710 \pm 0.05	26.80 \pm 0.09db

^aPSNR = peak signal-to-noise ratio; SSIM = structural similarity index measure.

In the above table, the ResUnet architecture is renowned for its capacity to alleviate the issue of gradient vanishing caused by increasing depth of layers, and has proven its proficiency in image reconstruction tasks. To further enhance ResUnet's effectiveness, we incorporated the

Convolutional Block Attention Module (CBAM) during the training phase. We refer to this network structure model as ResCBAM. The results from the SSIM and PSNR values in the Table 1 clearly indicate that FDP-GAN exhibits a remarkable proficiency in handling the fine details of these samples.

3.2.2. Numerical simulation

In this experiment, the reference image was obtained through reconstruction using 512 projections (Fig. 5(a)). A reduced number of sensors were used to reconstruct an image, which served as the artificial input for the network. Additionally, a numerical simulation was conducted to highlight the presence of streak-type artifacts in the image (Fig. 5(b)). The simulated image was then used as input to the network to evaluate the de-artifact visualization effect of the model, with the resulting output map presented in Fig. 5(c). Notably, the theoretical image was regarded as the reference, while the reconstructed image with a limited sensor configuration was utilized as an artificial input. The numerical simulations performed in this work are intended to illustrate the effects of streak-type artifacts in images.

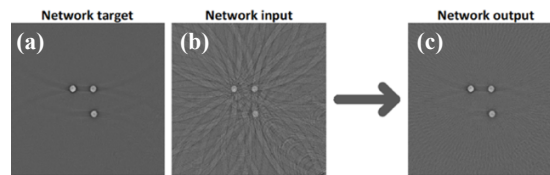


Fig. 5. Numerical simulation results. (a): Network Target: Numerical Simulation Results in 512 Dimensions; (b): Tomographic Reconstruction of a Single-Point Absorber: Numerical Simulation with 16 Detection Positions (Projections) Utilizing a 2D Filtered back projection (FBP) Algorithm; (c): Visualization of FDP-GAN Network Output Results.

3.2.3. Performance evaluation on PAT datasets

In processing PAT data, the number of sensors plays a crucial role in determining the performance and functionality of the device [27]. Particularly, in cases where the imaging process involves a reduced number of sensors, spatial resolution may suffer, resulting in a constrained imaging range. This can subsequently lead to the introduction of artifacts or under-sampling issues in the acquired images. Consequently, we employ deep learning techniques specifically to address the challenges posed by imaging at low sensor counts. We performed our experiments on abdomen of mice data under a variety of undersampling conditions 16, 32, 64, 128, and 256, where 512 was set as the label.

In Fig. 6, we conduct a comparative analysis of training performance among different models with varying sensor numbers. Notably, utilizing 16 projections, the FDP-GAN model demonstrates remarkable image contour and detail, akin to pix2pix with 32 projections. Furthermore, FDP-GAN exhibits image quality recovery approaching the ground truth value of 512 projections at 32 projections. The SSIM and PSNR metrics in Tables 2 and 3 underscore FDP-GAN's superior performance, particularly at lower projection rates of 16 and 8. Figure 7's radar, line, and box plots illustrate SSIM and PSNR metrics for the complete test set, validating the efficacy of each model's training.

3.2.4. Ablation experiment

In this study, ablation experiments were conducted to optimize and fine-tune the pre-training parameters in deep learning network architectures. This entailed a series of controlled manipulations where designated elements underwent selective alterations or removal to gauge their impact

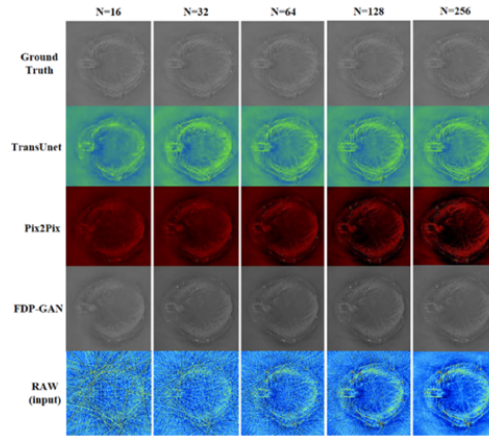


Fig. 6. The training results of different network structures were compared based on the number of transducers. The comparison was conducted using images of the rat brain with different transducer numbers, specifically $N = 8$, $N = 16$, $N = 32$, $N = 64$, and $N = 128$.

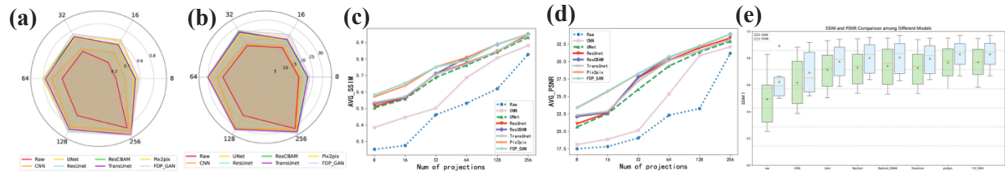


Fig. 7. PAT imaging qualitative analysis visualization results. (a)(c) Radar plot and discounted line plot presenting SSIM metrics for training results with diverse network structures and sensor counts. (b)(d) Radar plot and discounted line plot visualizing PSNR metrics for training outcomes with varying network structures and sensor counts. (e) Box line plots demonstrating the training results across different network structures and sensor counts.

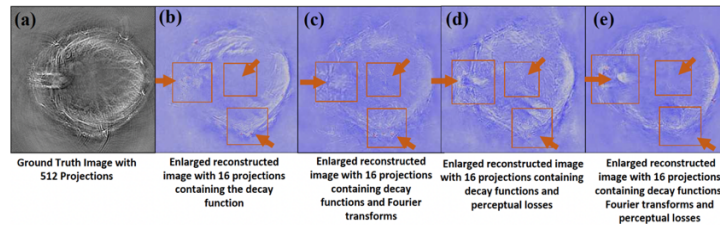


Fig. 8. Comparative Analysis of Ground Truth (512 Projections) and Recovered Image (16 Projections). SubFigures (a), (b), (c), (d), and (e) Showcase the Impact of Sequentially Adding Decay Function Loss, FFT-Loss, and Perceptual Loss, Respectively.

on the overall model efficacy. Within the framework of the FDP-GAN model, comprehensive ablation experiments were executed to probe the effects of the decay function, Fourier Transform, perceptual loss function, and L1 loss on the refinement of contour edge features. Through judicious activation or deactivation of these constituents, the investigation sought to elucidate their respective roles in optimizing image quality.

The ablation experiment results (Fig. 8) reveal the impact of incorporating decay function loss, fft-loss, and perceptual loss sequentially in the configuration with 16 transducers. Images (b), (c), (d), and (e) in Fig. 8 demonstrate the effects of each variable on different parts of the

image. Figure 8(a) illustrates the ground truth image as the label image under 512 transducers. Specifically, (b) enhances overall smoothness, (c) reveals clearer internal details, and (d) further enhances both internal details and external contours. In (e), after adding fft-loss to (d), brain slice details and external contours are further highlighted, approaching ground truth. Table 4 provides quantitative evaluation, demonstrating superior image enhancement with metrics such as PSNR and SSIM. Ablation experiments affirm the denoising and de-artifacting performance of our FDP-GAN network, preserving intricate high-frequency details.

4. Discussion

This research proposes a novel method for image processing using a Fourier perceptual loss generative adversarial network (FDP-GAN), which effectively blends a traditional physical iterative approach with GAN networks. This approach is rooted in a pix2pix GAN module and includes features such as Fourier transform, a decay function, and perceptual and L_1 losses. Our model's application to manually collected PAM and PAT datasets showcases exceptional performance in both quantitative and qualitative assessments. The amalgamation of optical and acoustic techniques in this integrated methodology holds significant promise for propelling advanced structural and functional imaging in pivotal biomedical fields like oncology, vascular biology, and neuroscience. However, FDP-GAN's effectiveness is diminished when dealing with highly undersampled imaging or anomalous data. Despite its successes, the method requires improvements to address challenges with image sampling below the count of 16 sensors and the influence of grayscale values in focal areas of simulated images on its performance. Also, the time consumed during the training of multiple modules is an area of concern.

The findings open doors to various possibilities in image processing. The network can be modified by choosing different constraints as regularization terms or by adopting different network structures for degraded image recovery in future iterations. Besides, the proposed architecture could potentially extend to other imaging techniques, like two-photon imaging [28,29] and second-harmonic imaging [30,31]. The FDP-GAN's potential goes beyond simple image refinement. The intersection of AI [32] and high-resolution [33–36] photoacoustic imaging could offer enhanced early disease detection capabilities and further accelerate the evolution of AI-driven medical applications.

5. Conclusion

This research presents an innovative deep-learning technique for reconstructing PAT from data with sparse sampling and finite views. By harnessing the GAN architecture and fine-tuning a regularized cost function, the method outperforms other deep learning techniques by providing superior image recovery, even in low-sampling scenarios (16 transducers). The proposed FDP-GAN is a pivotal advancement in this research, this model is designed to leverage deep features found both in the time and frequency domains of the original time-axis signal, using data-driven feature projection parameters.

Funding. Science and Technology Innovation Engineering Project of the Chinese Academy of Traditional Chinese Medicine; Climbing Program Foundation from Beijing Institute of Petrochemical Technology; Beijing Municipal Education Commission (22019821001).

Acknowledgments. The work is supported by the fund of the Beijing Municipal Education Commission, China, under grant number 22019821001, the fund of Climbing Program Foundation from Beijing Institute of Petrochemical Technology (Project No. BIPTAAI-2021007), and the fund of Science and Technology Innovation Engineering Project of the Chinese Academy of Traditional Chinese Medicine (Project No. CI2021A01910).

Disclosures. The authors declare no competing interests.

Data availability. PAM Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request. PAT Data underlying the results presented in this paper are available in Ref.[11].

References

1. R. Rai, V. N. Rai, and S. N. Thakur, "Photoacoustic tomography and its applications," in *Photoacoustic and Photothermal Spectroscopy* (Elsevier, 2023), pp. 621–645.
2. Y. Gu, Y. Sun, X. Wang, *et al.*, "Application of photoacoustic computed tomography in biomedical imaging: A literature review," *Bioeng. Transl. Med.* **8**, e10419 (2022).
3. Z. Wei, W. Liu, W. Yu, *et al.*, "Multiple Parallel Fusion Network for Predicting Protein Subcellular Localization from Stimulated Raman Scattering (SRS) Microscopy Images in Living Cells," *Int. J. Mol. Sci.* **23**(18), 10827 (2022).
4. S. Liu, H. Wang, C. Zhang, *et al.*, "In vivo photoacoustic sentinel lymph node imaging using clinically-approved carbon nanoparticles," *IEEE Trans. Biomed. Eng.* **67**, 2033 (2019).
5. N. Guezzi, C. Lee, T. D. Le, *et al.*, "Multistage adaptive noise reduction technique for optical resolution photoacoustic microscopy," *J. Biophotonics* **15**(12), e202200164 (2022).
6. L. Li, L. Zhu, C. Ma, *et al.*, "Single-impulse panoramic photoacoustic computed tomography of small-animal whole-body dynamics at high spatiotemporal resolution," *Nat. Biomed. Eng.* **1**(5), 0071 (2017).
7. S. Hakakzadeh, M. Amjadian, Y. Zhang, *et al.*, "Signal restoration algorithm for photoacoustic imaging systems," *Biomed. Opt. Express* **14**(2), 651 (2023).
8. T. Wang, W. Liu, and C. Tian, "Combating acoustic heterogeneity in photoacoustic computed tomography: A review," *J. Innov. Opt. Health Sci.* **13**(03), 2030007 (2020).
9. Y. Xu and L. V. Wang, "Effects of acoustic heterogeneity in breast thermoacoustic tomography," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.* **50**(9), 1134–1146 (2003).
10. N. Davoudi, X. L. Deán-Ben, and D. Razansky, "Deep learning optoacoustic tomography with sparse data," *Nat Mach Intell* **1**(10), 453–460 (2019).
11. A. Hauptmann, F. Lucka, M. Betcke, *et al.*, "Model-Based Learning for Accelerated, Limited-View 3-D Photoacoustic Tomography," *IEEE Trans. Med. Imaging* **37**(6), 1382–1393 (2018).
12. T. Di Ianni and R. D. Airan, "Deep-fUS: A Deep Learning Platform for Functional Ultrasound Imaging of the Brain Using Sparse Data," *IEEE Trans. Med. Imaging* **41**(7), 1813–1825 (2022).
13. S. Antholzer, M. Haltmeier, and J. Schwab, "Deep learning for photoacoustic tomography from sparse data," *Inverse Probl. Sci. Eng.* **27**(7), 987–1005 (2019).
14. Z. Zhao, B. Shen, Y. Li, *et al.*, "Deep learning-based high-speed, large-field, and high-resolution multiphoton imaging," *Biomed. Opt. Express* **14**(1), 65 (2023).
15. S. Guan, A. A. Khan, S. Sikdar, *et al.*, "Fully Dense UNet for 2-D Sparse Photoacoustic Tomography Artifact Removal," *IEEE J. Biomed. Health Inform.* **24**(2), 568–576 (2020).
16. M. L. Terranova, "Prominent Roles and Conflicted Attitudes of Eumelanin in the Living World," *Int. J. Mol. Sci.* **24**(9), 7783 (2023).
17. S. J. Sterbing-D'Angelo, H. Liu, M. Yu, *et al.*, "Morphology and deflection properties of bat wing sensory hairs: scanning electron microscopy, laser scanning vibrometry, and mechanics model," *Bioinspir. Biomim.* **11**(5), 056008 (2016).
18. Z. Zhang, Z. Chen, L. Shang, *et al.*, "Structural Color Materials from Natural Polymers," *Adv. Mater. Technol. (Weinheim, Ger.)* **6**(11), 2100296 (2021).
19. P. Freyer, B. D. Wilts, and D. G. Stavenga, "Cortex Thickness Is Key for the Colors of Iridescent Starling Feather Barbules With a Single, Organized Melanosome Layer," *Front. Ecol. Evol.* **9**, 746254 (2021).
20. M. Giraldo, J. Sosa, and D. Stavenga, "Feather iridescence of *Coeligena* hummingbird species varies due to differently organized barbules and barbules," *Biol. Lett.* **17**(8), 20210190 (2021).
21. A. Refaee, C. J. Kelly, H. Moradi, *et al.*, "Denoising of pre-beamformed photoacoustic data using generative adversarial networks," *Biomed. Opt. Express* **12**(10), 6184 (2021).
22. K. Nakamura and B.-W. Hong, "Adaptive weight decay for deep neural networks," *IEEE Access* **7**, 118857–118865 (2019).
23. R. S. Burt, "Decay functions," *Social networks* **22**(1), 1–28 (2000).
24. S. Huang, Y. He, S. Yu, *et al.*, "Experimental investigation and prediction model for UCS loss of unsaturated sandstones under freeze-thaw action," *Int. J. Min. Sci. Technol.* **32**(1), 41–49 (2022).
25. K. Zhang and Y.-X. Wang, "Deep Learning meets Nonparametric Regression: Are Weight-Decayed DNNs Locally Adaptive?" *arXiv*, arXiv preprint arXiv:2204.09664 (2022).
26. J. Johnson, A. Alahi, L. Fei-Fei, *et al.*, eds., *Lecture Notes in Computer Science* (Springer International Publishing, 2016), Vol. 9906, pp. 694–711.
27. X. Zhu, Q. Huang, A. DiSpirito, *et al.*, "Real-time whole-brain imaging of hemodynamics and oxygenation at micro-vessel resolution with ultrafast wide-field photoacoustic microscopy," *Light: Sci. Appl.* **11**(1), 138 (2022).
28. L. D. Faulhaber, O. D'Costa, A. Y. Shih, *et al.*, "Antibody-based in vivo leukocyte label for two-photon brain imaging in mice," *Neurophoton.* **9**(03), 031917 (2022).
29. T. Wang, M. He, K. Shen, *et al.*, "Learned regularization for image reconstruction in sparse-view photoacoustic tomography," *Biomed. Opt. Express* **13**(11), 5721 (2022).
30. M. Fouad, M. A. A. E. Ghany, and G. Schmitz, "A Single-Shot Harmonic Imaging Approach Utilizing Deep Learning for Medical Ultrasound," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.* **70**(3), 237–252 (2023).
31. Z. Wei, X. Wu, W. Tong, *et al.*, "Elimination of stripe artifacts in light sheet fluorescence microscopy using an attention-based residual neural network," *Biomed. Opt. Express* **13**(3), 1292–1311 (2022).

32. D. Seong, E. Lee, Y. Kim, *et al.*, “Three-dimensional reconstructing undersampled photoacoustic microscopy images using deep learning,” *Photoacoustics* **29**, 100429 (2023).
33. C. Qiao, D. Li, Y. Liu, *et al.*, “Rationalized deep learning super-resolution microscopy for sustained live imaging of rapid subcellular processes,” *Nat. Biotechnol.* **41**(3), 367–377 (2023).
34. J. Kim, G. Kim, L. Li, *et al.*, “Deep learning acceleration of multiscale superresolution localization photoacoustic imaging,” *Light: Sci. Appl.* **11**(1), 131 (2022).
35. M. Amjadian, S. M. Mostafavi, J. Chen, *et al.*, “Super-Resolution Photoacoustic Microscopy via Modified Phase Compounding,” *IEEE Trans. Med. Imaging* **41**(11), 3411–3420 (2022).
36. H. L. Chua and A. Huong, “Determination of Solution Concentrations using Photoacoustic Technology and Deep Learning: A comparison study,” in *2023 IEEE 13th Symposium on Computer Applications & Industrial Electronics (ISCAIE)* (IEEE, 2023), pp. 66–71.